Splitting methods for second-order initial value problems

P.J. van der Houwen, E. Messina

Modelling, Analysis and Simulation (MAS)

MAS-R9809 July 31, 1998

Report MAS-R9809
ISSN 1386-3703

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Splitting Methods for Second-Order Initial Value Problems

P.J. van der Houwen
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

&

E. Messina
Dipartimento di Matematica e Applicazioni "R. Caccioppoli"
University of Napels "Federico II", Via Cintia, I-80126 Naples, Italy

ABSTRACT
We consider stiff initial-value problems for second-order differential equations of the special form $\mathbf{y}'' = f(\mathbf{y})$. Stiff initial-value problem solvers are necessarily implicit, hence, we are faced with the problem of solving systems of implicit relations. This paper focuses on the construction and analysis of iterative solution methods which are effective in cases where the Jacobian of the righthand side of the differential equation can be split into a sum of matrices with a simple structure. These iterative methods consist of the modified Newton method and an iterative linear solver to deal with the linear Newton systems. The linear solver is based on the approximate factorization of the system matrix associated with the linear Newton systems. A number of convergence results are derived for the linear solver in the case where the Jacobian matrix can be split into commuting matrices. Such often problems arise in the spatial discretization of time-dependent partial differential equations. Furthermore, the stability matrix and the order of accuracy of the integration process are derived in the case of a *finite* number of iterations.

1991 Mathematics Subject Classification: 65L06
Keywords and Phrases: Second-order partial differential equations, splitting methods, approximate factorization.
Note: Work carried out under project MAS 1.4 - Exploratory research: Analysis of ODEs and PDEs.

1. Introduction

We consider initial-value problems (IVPs) for systems of second-order ordinary differential equations (ODEs) of the special form

$$(1.1)\quad \frac{d^2\mathbf{y}(t)}{dt^2} = \mathbf{f}(\mathbf{y}(t)), \qquad \mathbf{y}, \mathbf{f} \in \mathbb{R}^d.$$

We shall assume that the equation (1.1) is stiff, so that we need a stiff solver to integrate (1.1). Stiff IVP solvers are necessarily implicit, hence, we are faced with the problem of solving systems of implicit relations. This paper focuses on the construction and analysis of iterative solution methods which are effective in cases where an approximation J to $\partial\mathbf{f}/\partial\mathbf{y}$ can be split into a sum of σ matrices J_i such that the matrices J_i have an essentially simpler structure than the matrix J (in Section 3.2, we will

specify what is meant by an 'essentially simpler structure'). These iterative methods consist of the modified Newton method (the outer iteration), in which the linear Newton systems are solved by a second iteration process (the inner iteration) which is based on approximate factorization. The inner-outer iteration process will be called *approximate factorization iteration* or briefly AF iteration.

In [5] AF iteration was used for solving fully implicit discretizations of transport models and in [2] AF iteration was analysed in the case of a large class of implicit integration methods for systems of first-order ODEs originating from the semidiscretization of partial differential equations. In the latter paper, general convergence and stability results are presented. These results can also be used for second-order ODE methods by writing (1.1) as a first-order system and by simply integrating this system by a first-order ODE solver (the black box approach). Unfortunately, in the usual case where the eigenvalues of $\partial\mathbf{f}/\partial\mathbf{y}$ are negative, the convergence and stability properties of the black box approach are quite poor, because the special structure of the first-order form of (1.1) is not exploited. To illustrate this, consider a Runge-Kutta (RK) method for first-order ODEs $\mathbf{y}' = \mathbf{g}(\mathbf{y})$, let the Butcher matrix $\tilde{A}$ of the RK method be an arbitrary matrix with *complex* eigenvalues, and suppose that $\partial\mathbf{g}/\partial\mathbf{y}$ can be written as the sum of two commuting matrices K_1 and K_2. Then it can be shown that the approximate factorization iteration process cannot be unconditionally convergent if the eigenvalues of K_1 and K_2 are purely imaginary (see [2]). Now we apply the same RK method to the first-order form of (1.1). Suppose that the Jacobian associated with the righthand side of (1.1) can be split into two matrices J_1 and J_2 which share the same eigensystem with negative eigenvalues (for example, this happens if (1.1) originates from the spatial discretization of a two-dimensional wave equation). Then, the matrices K_1 and K_2 associated with the first-order form $\mathbf{y}' = \mathbf{g}(\mathbf{y})$ of (1.1) commute and their eigenvalues are purely imaginary. Hence, according to [2], the AF iteration process for solving the implicit RK relations will not be unconditionally convergent. However, exploiting the special structure of the first-order form $\mathbf{y}' = \mathbf{g}(\mathbf{y})$ of (1.1), the implicit RK relations can be simplified (see Section 2 for details) and applying AF iteration to these simplified relations, we obtain unconditional convergence provided that the eigenvalues $\lambda(\tilde{A})$ of the underlying Butcher matrix $\tilde{A}$ satisfy $|\arg(\lambda(\tilde{A}))| \le \pi/4$. Examples are the Butcher matrices of the third-order Radau IIA, the fourth-order Lobatto IIIA, and the fourth-order and sixth-order Gauss methods. Thus, although the *solutions* of the original and the simplified RK relations are identical, the *convergence properties* of AF iteration are quite different.

The purpose of this paper is to see to what extent the convergence and stability results valid for first-order ODE methods change in the second-order case (1.1). Our starting point is the class of so-called General Linear Methods (GLMs). For first-order ODEs, such methods have been introduced by Butcher in 1966 (see [1, p.335] for a detailed discussion). In Section 2, we show that GLM methods can be defined in a similar way for second-order ODEs given by (1.1). The advantage of using the GLM format is that almost any IVP solver can be written as a GLM, so that the analysis developed in this paper applies to a wide variety of methods. Section 3 discusses the structure of the implicit relations arising in these GLM and defines the outer-inner iteration process for the implicit stage values. In Section 4, a number of convergence results are derived for the model situation where the

matrices J_i share the same eigensystem and possess a negative eigenvalue spectrum. Finally, Section 5 presents order of accuracy and stability results in the case of a *finite* number of inner and outer iterations.

2. General linear methods

A direct extension of the GLMs of Butcher to equations of the second-order form (1.1) reads

$$\mathbf{U}_{n+1} = (R\otimes I)\mathbf{U}_n + h^2(S\otimes I)\mathbf{F}(\mathbf{U}_n) + h^2(T\otimes I)\mathbf{F}(\mathbf{U}_{n+1}), \quad n = 1, 2, \ldots . \tag{2.1}$$

Here R, S and T denote k-by-k matrices, I is the d-by-d identity matrix, h is the stepsize $t_{n+1} - t_n$, and $\otimes$ denotes the Kronecker product, i.e. if $R = (r_{ij})$, then $R\otimes I$ denotes the matrix of matrices $(r_{ij}I)$. In this paper, we assume that each of the k components $\mathbf{u}_{n+1,i}$ of the kd-dimensional solution vector $\mathbf{U}_{n+1}$ represents a numerical approximation either to the exact solution vector $\mathbf{y}(t_n + a_ih)$ or to the exact derivative vector $h\mathbf{y}'(t_n + a_ih)$. The vector $\mathbf{a} := (a_i)$ is called the *abscissa vector*, the quantities $\mathbf{U}_{n+1}$ the *stage vectors* and their components $\mathbf{u}_{n+1,i}$ the *stage values*. The stage values approximating $\mathbf{y}(t_n + a_ih)$ will be called *solution values* and those approximating $h\mathbf{y}'(t_n + a_ih)$ *derivative values*. Furthermore, for any vector $\mathbf{U}_n = (\mathbf{u}_{ni})$, $\mathbf{F}(\mathbf{U}_n)$ contains the righthand side values $(\mathbf{f}(\mathbf{u}_{ni}))$.

If in the formula (2.1) h^2 is replaced by h, then we obtain a GLM for first-order ODEs. In both cases, the GLM is completely determined by the arrays $\{\mathbf{a}, R, S, T\}$. Given the starting vector $\mathbf{U}_1$, (2.1) defines a sequence of vectors $\mathbf{U}_2$, $\mathbf{U}_3$, $\mathbf{U}_4$, ..., from which approximations to the exact solution values can be obtained.

It may happen that R and S have zero columns for the same column index j. In such cases, the jth component $\mathbf{u}_{1,j}$ of $\mathbf{U}_1$ is not needed to start the integration process. All stage values that we do need to start the method are called *external* stage values, otherwise they are called *internal* stage values (cf. Butcher [1], p. 367). The distinction between internal and external stage values is needed in the stability analysis given in Section 5.

In this paper, we shall assume that one or more abscissae a_i equal 1. If the corresponding components $\mathbf{u}_{n+1,i}$ of $\mathbf{U}_{n+1}$ are external stage values, then these components will be called *step point values* (the points t_{n+1} are called *step points*). A stage value $\mathbf{u}_{n+1,i}$ which provides an approximation to the exact solution value $\mathbf{y}(t_n + a_ih)$ is said to be accurate of order p if for sufficiently smooth righthand side functions $\mathbf{f}$ and for all points $\{t_n + a_ih, n = 0, 1, \ldots\}$, we have that $\mathbf{u}_{n+1,i} = \mathbf{y}(t_n + a_ih) + O(h^p)$. The *maximal* order of accuracy of the step point values is called the *step point order*.

Of course, the second-order ODE (1.1) can also be solved by reducing the ODE (1.1) to first-order form and by application of a first-order-ODE method. There are now two options, (i) the *black box* approach where the first-order-ODE method is used as a black box method, or (ii) the *indirect* second-order-ODE-method approach where the first-order-ODE method is rewritten as a second-order ODE method by exploiting the special structure of the first-order ODE system. In the black box option, we have to rely on the properties of the first-order-ODE method, including the properties of the iteration process implemented for solving the implicit relations. Since it is often more advantageous, with respect to numerical performance, to follow the indirect second-order-ODE-method option, we

explicitly derive the resulting second-order ODE method. Let us write (1.1) as $\mathbf{y}' = \mathbf{z}$, $\mathbf{z}' = \mathbf{f}(\mathbf{y})$ and let us apply a GLM defined by the arrays $(\tilde{\mathbf{a}}, \tilde{R}, \tilde{S}, \tilde{T})$. It can be verified that the resulting method is equivalent with separately applying this GLM to $\mathbf{y}' = \mathbf{z}$ and to $\mathbf{z}' = \mathbf{f}(\mathbf{y})$. Hence, let us associate with $\mathbf{y}$ and $\mathbf{z}$ the stage vectors $\mathbf{Y}$ and $\mathbf{Z}$. Then, $\mathbf{Y}$ and $\mathbf{Z}$ satisfy

$$\mathbf{Y}_{n+1} = (\tilde{R} \otimes I)\mathbf{Y}_n + h(\tilde{S} \otimes I)\mathbf{Z}_n + h(\tilde{T} \otimes I)\mathbf{Z}_{n+1},$$
$$\mathbf{Z}_{n+1} = (\tilde{R} \otimes I)\mathbf{Z}_n + h(\tilde{S} \otimes I)\mathbf{F}(\mathbf{Y}_n) + h(\tilde{T} \otimes I)\mathbf{F}(\mathbf{Y}_{n+1}).$$

By substitution of the second equation into the first and by defining the extended stage vector $\mathbf{U}_n := (\mathbf{Y}_n^T, h\mathbf{Z}_n^T)^T$, we obtain a GLM for second-order ODEs (see also Hairer [3])

$$(2.2) \qquad \mathbf{a} = \begin{pmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{a}} \end{pmatrix}, \quad R = \begin{pmatrix} \tilde{R} & \tilde{S} + \tilde{T}\tilde{R} \\ O & \tilde{R} \end{pmatrix}, \quad S = \begin{pmatrix} \tilde{T}\tilde{S} & O \\ \tilde{S} & O \end{pmatrix}, \quad T = \begin{pmatrix} \tilde{T}^2 & O \\ \tilde{T} & O \end{pmatrix}.$$

Note that in (2.2) only $\mathbf{Y}_{n+1}$ is implicitly defined and should be solved by some iteration process. Thus, this iteration process needs to be applied to only kd implicit relations. This is a direct consequence of the special structure of the first-order system. Ignoring this special structure, that is, applying the black box option (i), would lead to iteration of 2kd implicit relations. Of course, if the iteration processes used in the two options both converge, then they converge to the same numerical solution. However, it will turn out that the iteration process in the indirect second-order-ODE-method approach often converges where it does not converge in the black box approach.

Example 2.1. An example of a GLM of the form (2.2) with step point order p = 2 is the GLM

$$(2.3) \qquad \mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad R = \frac{1}{9}\begin{pmatrix} 0 & 9 & 0 & 0 \\ -3 & 12 & -2 & 8 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & -3 & 12 \end{pmatrix}, \quad S = O, \quad T = \frac{1}{9}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \end{pmatrix}.$$

derived from the two-step backward differentiation method (BDM). Here, $\mathbf{U}_{n+1}$ approximates $(\mathbf{y}(t_n)^T, \mathbf{y}(t_n + h)^T, h\mathbf{y}'(t_n)^T, h\mathbf{y}'(t_n + h)^T)^T$. ♦

Example 2.2. Another *indirect* second-order ODE method, derived from the 2-stage Radau IIA based method for first-order ODEs, is defined by the Runge-Kutta-Nyström (RKN) method

$$(2.4) \qquad \mathbf{a} = \begin{pmatrix} 1/3 \\ 1 \\ 1 \end{pmatrix}, \quad R = \frac{1}{6}\begin{pmatrix} 0 & 6 & 2 \\ 0 & 6 & 6 \\ 0 & 0 & 6 \end{pmatrix}, \quad S = O, \quad T = \frac{1}{72}\begin{pmatrix} 8 & -4 & 0 \\ 36 & 0 & 0 \\ 54 & 18 & 0 \end{pmatrix},$$

where $\mathbf{U}_{n+1} \approx (\mathbf{y}(t_n + h/3)^T, \mathbf{y}(t_n + h)^T, h\mathbf{y}'(t_n + h)^T)^T$. This method has step point order 3. ♦

Example 2.3. A *direct* second-order ODE method is given by (cf. Sharp, Fine and Burrage [7])

$$(2.5)\qquad \mathbf{a} = \begin{pmatrix} 17/14 \\ 23/60 \\ 1 \\ 1 \end{pmatrix},\quad R = \begin{pmatrix} 0 & 0 & 1 & 17/14 \\ 0 & 0 & 1 & 23/60 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},\quad S = O,\quad T = \begin{pmatrix} 289/392 & 0 & 0 & 0 \\ -234179/352800 & 289/392 & 0 & 0 \\ -21/698 & 185/349 & 0 & 0 \\ 49/349 & 300/349 & 0 & 0 \end{pmatrix},$$

where $\mathbf{U}_{n+1}$ approximates $\big(\mathbf{y}(t_n + 17h/14)^T,\ \mathbf{y}(t_n + 23h/60)^T,\ h\mathbf{y}(t_n + h)^T,\ h\mathbf{y}'(t_n + h)^T\big)^T$. This method has step point order 3. ♦

3. Approximate factorization iteration

In order to define the approximate factorization iteration method, we first need to extract the implicit relations to be solved from the GLM (2.1). This will be the subject of Section 3.1. In Section 3.2, we will specify the iteration method by using the splitting mentioned in the introduction.

3.1. Structure of the implicit relations

To see the structure of the implicit relations to be solved, it is convenient to partition the components $\mathbf{u}_{n+1,i}$ of $\mathbf{U}_{n+1}$ into (i) *explicit* stage values that can be explicitly evaluated by means of already computed stage values and righthand side values, and (ii) *implicit* stage values which need the solution of a (usually nonlinear) system of equations. For instance, in example (2.3), all stage values are explicit except for the second one, and in (2.4) and (2.5), only the first two stages are implicit and the other stages are explicit.

In most methods available in the literature, the components of $\mathbf{U}_{n+1}$ can be arranged in such a way that $\mathbf{U}_{n+1} = (\mathbf{X}_{n+1}{}^T,\ \mathbf{Y}_{n+1}{}^T,\ \mathbf{Z}_{n+1}{}^T)^T$, where $\mathbf{X}_{n+1}$ and $\mathbf{Z}_{n+1}$ represent explicit stage values and $\mathbf{Y}_{n+1}$ the implicit stage values (see again the examples (2.3), (2.4) and (2.5)). The corresponding partitioning of the matrix T takes the form

$$(3.1)\qquad T = \begin{pmatrix} L_1 & O & O \\ T_{21} & A & O \\ T_{31} & T_{32} & L_2 \end{pmatrix},$$

where L_1 and L_2 are strictly lower triangular matrices and T_{21}, T_{31}, T_{32} and A are allowed to be full matrices with A nonsingular. From (3.1) it follows that the implicit stage values are defined by

$$(3.2)\qquad \mathbf{R}_n(\mathbf{Y}_{n+1}) = \mathbf{0},\quad \mathbf{R}_n(\mathbf{Y}) := \mathbf{Y} - h^2(A\otimes I)\mathbf{F}(\mathbf{Y}) - \mathbf{V}_n,$$

where $\mathbf{V}_n$ can be expressed in terms of already computed quantities. The structure of the implicit relations defining the implicit stage values is mainly determined by the matrix A. For the implicit GLMs defined by (2.2), (2.3), (2.4) and (2.5), the matrix A is respectively given by

$$(3.3)\qquad A = \widetilde{T}^{\,2},\quad A = \frac{4}{9},\quad A = \frac{1}{72}\begin{pmatrix} 8 & -4 \\ 36 & 0 \end{pmatrix},\quad A = \begin{pmatrix} 289/392 & 0 \\ -234179/352800 & 289/392 \end{pmatrix},$$

where we assumed that in (2.2) the matrix $\tilde{T}$ is nonsingular. In the following, the number of implicit stages will be denoted by s.

Before discussing the solution of the implicit relation (3.2), we remark that for stiff problems it is recommendable to impose a special structure on the matrices S and T such that the evaluation of *explicit* righthand side values can be avoided. This considerably improves the accuracy in actual implementations. To be more precise, let R, S and T be partitioned according to the partitioning $\mathbf{U}_{n+1} = (\mathbf{X}_{n+1}^T, \mathbf{Y}_{n+1}^T, \mathbf{Z}_{n+1}^T)^T$, and let

$$(3.4) \qquad R = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}, \quad S = \begin{pmatrix} O & S_{12} & O \\ O & S_{22} & O \\ O & S_{32} & O \end{pmatrix}, \quad T = \begin{pmatrix} O & O & O \\ O & A & O \\ O & T_{32} & O \end{pmatrix}$$

where A is a nonsingular s-by-s matrix (note that the methods (2.3), (2.4) and (2.5) possess parameter matrices of this form). The GLM takes the form

$$\mathbf{X}_{n+1} = (R_1 \otimes I)\mathbf{U}_n + h^2(S_{12} \otimes I)\mathbf{F}(\mathbf{Y}_n),$$

$$\mathbf{Y}_{n+1} = (R_2 \otimes I)\mathbf{U}_n + h^2(S_{22} \otimes I)\mathbf{F}(\mathbf{Y}_n) + h^2(A \otimes I)\mathbf{F}(\mathbf{Y}_{n+1}),$$

$$\mathbf{Z}_{n+1} = (R_3 \otimes I)\mathbf{U}_n + h^2(S_{32} \otimes I)\mathbf{F}(\mathbf{Y}_n) + h^2(T_{32} \otimes I)\mathbf{F}(\mathbf{Y}_{n+1}).$$

Using a similar approach as used by Shampine [6] in the implementation of implicit RK methods (see also Hairer and Wanner [4, p.129]), we express $\mathbf{F}(\mathbf{Y}_{n+1})$ in terms of $\mathbf{Y}_{n+1}$, $\mathbf{U}_n$ and $\mathbf{F}(\mathbf{Y}_n)$, i.e.

$$(3.5a) \qquad h^2\mathbf{F}(\mathbf{Y}_{n+1}) = (A^{-1} \otimes I)\mathbf{Y}_{n+1} - (A^{-1}R_2 \otimes I)\mathbf{U}_n - h^2(A^{-1}S_{22} \otimes I)\mathbf{F}(\mathbf{Y}_n),$$

so that we can write the GLM in the equivalent form

$$\mathbf{X}_{n+1} = (R_1 \otimes I)\mathbf{U}_n + h^2(S_{12} \otimes I)\mathbf{F}(\mathbf{Y}_n),$$

$$(3.5b) \qquad \mathbf{Y}_{n+1} = (R_2 \otimes I)\mathbf{U}_n + h^2(S_{22} \otimes I)\mathbf{F}(\mathbf{Y}_n) + h^2(A \otimes I)\mathbf{F}(\mathbf{Y}_{n+1}),$$

$$\mathbf{Z}_{n+1} = \big((R_3 - T_{32}A^{-1}R_2) \otimes I\big)\mathbf{U}_n + h^2\big((S_{32} - T_{32}A^{-1}S_{22}) \otimes I\big)\mathbf{F}(\mathbf{Y}_n) + (T_{32}A^{-1} \otimes I)\mathbf{Y}_{n+1}.$$

Since $h^2\mathbf{F}(\mathbf{Y}_n)$ can be generated by applying (3.5a) for n = 1, 2, ... , n-1, no *explicit* **F** evaluations are needed in (3.5) except for $\mathbf{F}(\mathbf{Y}_1)$. We shall use the formulas (3.5b) in the stability analysis of the iterated GLM (see Section (5.2).

3.2. The iteration method

Each step by the method (2.1) requires the solution of the nonlinear system $\mathbf{R}_n(\mathbf{Y}) = \mathbf{0}$ specified in (3.2). In order to solve this system, we consider the modified Newton iteration process:

$$(3.6a) \qquad M\big(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}\big) = -\mathbf{R}_n(\mathbf{Y}^{(j-1)}), \quad M := I - A \otimes h^2J, \quad j = 1, 2, \ldots, m,$$

where M represents an approximation to the Jacobian matrix of $\mathbf{R}_n(\mathbf{Y})$. If the dimension d in the system (1.1) is large, then solving (3.6a) usually is quite costly. It is the aim of this paper to reduce these costs by designing a parallel iterative linear system solver based on a splitting of the Jacobian of **f** in a sum of matrices J_i. Then, the matrix M can be expressed as

$$(3.6b) \qquad M = I - A\otimes h^2J = \frac{1}{\sigma}\sum_{i=1}^{\sigma}\left(I - \sigma A\otimes h^2J_i\right).$$

This linear solver may be considered as the *inner* iteration process and the Newton process (3.6a) as the *outer* iteration process. The inner-outer iteration process analysed in this paper is based on the approximate factorization of the matrix M and is of the form

$$(3.7) \qquad \Pi\left(\mathbf{Y}^{(j,\nu)} - \mathbf{Y}^{(j,\nu-1)}\right) = M\left(\mathbf{Y}^{(j-1,r)} - \mathbf{Y}^{(j,\nu-1)}\right) - \mathbf{R}_n\left(\mathbf{Y}^{(j-1,r)}\right), \quad \Pi := \prod_{i=\sigma}^{1}\left(I - B\otimes h^2J_i\right),$$

where $\nu = 1, 2, \ldots, r$, $j = 1, 2, \ldots, m$, and where B is a suitably chosen matrix. Evidently, if the iterates $\mathbf{Y}^{(j,\nu)}$ converge with ν, then they can only converge to the solution of (3.6a) with $\mathbf{Y}^{(j)}$ replaced by $\mathbf{Y}^{(j-1,r)}$. We will refer to (3.7) as AF iteration.

Each inner iteration in (3.7) requires the solution of σ linear systems with system matrix $I - B\otimes h^2J_i$ of order sd. It is now clear what we meant by the preposition that the 'partial' Jacobians J_i should have an 'essentially simpler structure', viz. 'the solution of the linear systems with system matrix $I - B\otimes h^2J_i$ should be much more easy than solving the linear system in (3.6a)'.

The inner iteration process in (3.7) is particularly attractive if parallel computer systems are available, because the σ LU-decompositions of the system matrices $I - B\otimes h^2J_i$ can all be done concurrently. Moreover, if B is diagonal, then the factor matrices $I - B\otimes h^2J_i$ of the system matrix Π are block-diagonal, which enables us to decouple each of the linear systems into s subsystems which can again be solved concurrently. If B is not diagonal, but similar to a diagonal matrix with *real* diagonal entries, then we can diagonalize the iteration method (3.7) by means of a Butcher transformation $\mathbf{Y}^{(j,\nu)} = (Q\otimes I)\tilde{\mathbf{Y}}^{(j,\nu)}$, where Q is such that $D := Q^{-1}AQ$ is diagonal (see e.g. [4. 128]) Thus,

$$(3.7') \qquad \begin{aligned} \tilde{\Pi}\left(\tilde{\mathbf{Y}}^{(j,\nu)} - \tilde{\mathbf{Y}}^{(j,\nu-1)}\right) &= -(Q^{-1}\otimes I)M(Q\otimes I)\tilde{\mathbf{Y}}^{(j,\nu-1)} + (Q^{-1}\otimes I)\left(M\mathbf{Y}^{(j-1,r)} - \mathbf{R}_n\left(\mathbf{Y}^{(j-1,r)}\right)\right), \\ \tilde{\Pi} &:= (Q^{-1}\otimes I)\Pi(Q\otimes I) = \prod_{i=\sigma}^{1}\left(I - D\otimes h^2J_i\right). \end{aligned}$$

Evidently, the factor matrices $I - D\otimes h^2J_i$ of the system matrix $\tilde{\Pi}$ are again block-diagonal, allowing the same amount of parallelism as in the case where B is diagonal.

Before turning to the convergence properties of AF iteration, we remark that an important class of problems that can be effectively dealt with by the approach described above are the initial-value problems originating from the spatial discretization of wave equations of the form

$$\frac{\partial^2 u}{\partial t^2} = g\left(\frac{\partial^2 u}{\partial x_1^2} , \dots , \frac{\partial^2 u}{\partial x_\sigma^2} , x_1, \dots , x_\sigma \right),$$

Then, the splitting of the corresponding Jacobian yields matrices J_i which each correspond with a one-dimensional differential operator. Hence, solving the linear subsystems is relatively cheap. ♦

4. Convergence results

Let us consider the behaviour of the iteration error $\varepsilon^{(j,\nu)} := \mathbf{Y}^{(j,\nu)} - \mathbf{Y}_{n+1}$. From (3.2) and (3.7) it follows that

$$\begin{aligned} &\varepsilon^{(j,\nu)} = Z\, \varepsilon^{(j,\nu-1)} + h^2 \Pi^{-1}(A\otimes I)\mathbf{G}_n(\varepsilon^{(j-1,r)}), \quad Z := I - \Pi^{-1}M, \\ &\mathbf{G}_n(\varepsilon) := \mathbf{F}(\mathbf{Y}_{n+1} + \varepsilon) - \mathbf{F}(\mathbf{Y}_{n+1}) - (I\otimes J)\varepsilon, \end{aligned} \tag{4.1}$$

where J is the same approximation to the Jacobian matrix as used in (3.6) and Z represents the inner amplification matrix. After r inner iterations, this recursion yields

$$\varepsilon^{(j,r)} = Z^r \varepsilon^{(j-1,r)} + h^2 \left(I - Z^r\right) M^{-1}(A\otimes I)\mathbf{G}_n\left(\varepsilon^{(j-1,r)}\right), \tag{4.2}$$

where we assumed that $\mathbf{Y}^{(j,0)} = \mathbf{Y}^{(j-1,r)}$, i.e. $\varepsilon^{(j,0)} = \varepsilon^{(j-1,r)}$. Let $\mathbf{G}_n$ possess a Lipschitz constant $L_n(h)$ in the neighbourhood of the origin (with respect to the norm $\|.\|$) and let $L_n(h) = O(h^u)$, where u depends on the update strategy used in the evaluation of the Jacobian J (if J is updated every few steps, then u = 1). Furthermore, it is easily verified that $Z = (A - B)\otimes h^2 J + O(h^4)$, so that $Z^r = O(h^{\theta r})$, where $\theta = 2$ if $A \neq B$ and $\theta = 4$ if $A = B$. Hence, it follows from (4.2) that

$$\| \varepsilon^{(j,r)} \| \leq \left(O(h^{\theta r}) + O(h^{u+2}) \right) \| \varepsilon^{(j-1,r)} \| , \quad j \geq 1. \tag{4.2'}$$

This estimate shows that we at least have fast convergence of the nonstiff components. For example, if u = 1, then in each outer iteration the iteration error is damped by a factor $O(h^{\theta r}) + O(h^3)$. Hence, choosing $r = 4\theta^{-1}$, we may expect a convergence rate comparable with that of modified Newton.
So far, all our considerations were independent of the splitting of the Jacobian J. However, in the remainder of this section, we will focus on the convergence in the case of model problems.

4.1. The model problem

The case where the 'partial' Jacobians J_i all commute with each other, that is, they share the same eigensystem, will be referred to as the *model problem*. Such model situations occur if (1.1) originates from certain classes of second-order partial differential equations, such as the wave equation mentioned above.
For briefness of notation, we introduce the following convention. Let $E(h^2J_1, \dots , h^2J_\sigma)$ be a matrix depending on $h^2J_1, \dots , h^2J_\sigma$. Then the s-by-s matrix obtained by replacing the matrices h^2J_i by the

scalars z_i is denoted by $E(\mathbf{z})$, where $\mathbf{z} = (z_1, \ldots, z_\sigma)$. Thus, with the matrices M defined in (3.6b), Π defined in (3.7), and Z defined in (4.2) we associate the matrices

$$(4.3)\quad Z(\mathbf{z}) := I - \Pi^{-1}(\mathbf{z})M(\mathbf{z}), \quad M(\mathbf{z}) = I - (\mathbf{e}^T\mathbf{z})A, \quad \Pi(\mathbf{z}) = \prod_{i=\sigma}^{1} \left(I - z_i B\right),$$

where $\mathbf{e}$ is the σ-dimensional vector with unit entries. Evidently, if we choose $z_i := \lambda(J_i)h^2$ where $\lambda(J_i)$ denotes an eigenvalue of J_i, then in the case of the model problem defined above, the eigenvalues of the amplification matrix in (4.1) are given by those of the matrix $Z(\mathbf{z})$. The region of convergence can then be defined by the region in the $\mathbf{z}$-plane where $Z(\mathbf{z})$ has its eigenvalues $\zeta(\mathbf{z})$ within the unit circle. Assuming that the eigenvalues of the 'partial' Jacobians J_i are on the *negative real axis* (as is the case in many wave equation problems), we shall call the iteration method (3.7) *A(0)-convergent* if the region of convergence contains the region $\{\mathbf{z}: z_i \le 0\}$. The eigenvalues $\zeta(\mathbf{z})$, will be called the *amplification factors* of the inner iteration method.

4.2. Matrices B = A with real eigenvalues

We consider the convergence region of (3.7) in the case where B = A with $\lambda(A)$ real (for example, as in the methods (2.3) and (2.5)). The amplification factors are given by

$$(4.4)\qquad \zeta(\mathbf{z}) := 1 - \pi^{-1}(\mathbf{z})\mu(\mathbf{z}), \quad \mu(\mathbf{z}) := 1 - \lambda(A)(\mathbf{e}^T\mathbf{z}), \quad \pi(\mathbf{z}) := \prod_{i=1}^{\sigma} (1 - \lambda(A)z_i),$$

where $\lambda(A)$ denotes an eigenvalue of A. Let $\lambda(A) \ge 0$. Then it follows from (4.4) that A(0)-convergence is achieved if $2\pi(\mathbf{z}) - \mu(\mathbf{z}) > 0$ for $z_i \le 0$. Since we may write

$$\pi(\mathbf{z}) = \mu(\mathbf{z}) + p_2\lambda^2(A) + p_3\lambda^3(A) + \ldots + p_\sigma\lambda^\sigma(A),$$

where the coefficients p_i are nonnegative whenever $z_i \le 0$, we see that for $\lambda(A) \ge 0$ and $z_i \le 0$

$$2\pi(\mathbf{z}) - \mu(\mathbf{z}) = \mu(\mathbf{z}) + 2\left(p_2\lambda^2(A) + p_3\lambda^3(A) + \ldots + p_\sigma\lambda^\sigma(A)\right) > 0.$$

Theorem 4.1. If $\lambda(A) \ge 0$, then AF iteration {(3.7), B = A} is A(0)-convergent for all σ.♦

4.3. Matrices B = A with complex eigenvalues

If B = A with A having complex eigenvalues, then the convergence analysis is more complicated. We separately discuss the cases of two and three splitting terms ($\sigma = 2$ and $\sigma = 3$).

4.3.1. Two splitting terms. If $\sigma = 2$, then the amplification factor can be factorized according to

$$(4.5)\qquad \zeta(\mathbf{z}) = \lambda(A)z_1\left(1 - \lambda(A)z_1\right)^{-1}\lambda(A)z_2\left(1 - \lambda(A)z_2\right)^{-1}.$$

By requiring that the magnitude of both factors is less than 1, we see that for $\sigma = 2$ the region of convergence of the inner iteration method in (3.7) contains the domain

$$\mathbb{D} := \bigcap_{\lambda(A)} \{\mathbf{z}:\ z_j\, \mathrm{Re}(\lambda(A)) < \tfrac{1}{2}\ ,\ j = 1, 2\}.$$

Theorem 4.2. If $\mathrm{Re}(\lambda(A)) \geq 0$, then AF iteration {(3.7), B = A} is A(0)-convergent for $\sigma = 2$. ♦

Thus, AF iteration applied to the examples (2.3), (2.4) and (2.5) is A(0)-convergent. In the particular case of the indirect GLM (2.2), we immediately have by virtue of Theorem 4.2 the result:

Corollary 4.1. If the generating GLM $(\tilde{\mathbf{a}}\ ,\ \tilde{R}\ ,\ \tilde{S}\ ,\ \tilde{T}\ )$ in the indirect GLM (2.2) satisfies $|\arg(\lambda(\tilde{T}))| \leq \pi/4$, then AF iteration {(3.7), $A = B = \tilde{T}^{\,2}$} is A(0)-convergent for $\sigma = 2$. ♦

This corollary implies that for all indirect RKN methods generated by RK matrices whose Butcher matrices $\tilde{A}$ have their eigenvalues in the wedge $|\arg(\lambda(\tilde{A}))| \leq \pi/4$ AF iteration is A(0)-convergent. For example, this happens in the case of the third-order Radau IIA, the fourth-order Lobatto IIIA and the fourth-order and sixth-order Gauss methods.

Next, consider the case where A has eigenvalues with $\mathrm{Re}(\lambda(A)) < 0$, so that A(0)-convergence is not possible. In fact, the region of convergence consists of two strips along the negative z_1-axis and the negative z_2-axis. The plot in Figure 1 is typical for the form of the region of divergence in the third quadrant of the (z_1,z_2)-plane obtained for methods with $\mathrm{Re}(\lambda(A)) < 0$ (black part indicates divergence). Note that the convergence region is symmetric with respect to the line $z_1 = z_2$.

In a number of important applications, we do not need A(0)-convergence with respect to both z_1 and z_2. For example, in the 2-dimensional modeling of the water elevation in a river, we encounter a wave equation in which the resolution of the coordinate perpendicular to the river should be an order of magnitude smaller than the resolution of the coordinate along the river. Hence, the "stiffness" of the Newton systems (3.6a) comes from the direction perpendicular to the river, so that we need only unconditional convergence with respect to this direction. In such cases, a region of convergence as in Figure 1 is quite sufficient.

If we have stiffness with respect to both z_1 and z_2, then we should look at the disk, centered at the origin, which is contained in the region of convergence. From Figure 1 it follows that the radius of this disk can be determined by setting $z_1 = z_2$ on the boundary of the convergence region. Hence, the point $z_0\mathbf{e}$ is on the boundary of this convergence disk if z_0 is a solution (nearest to the origin) of the equations $|\zeta(z_0\mathbf{e})| = 1$ associated with those eigenvalues $\lambda(A)$ of A that are in the negative halfplane. From (4.5) it follows that z_0 satisfies $|\lambda(A)z_0(1 - \lambda(A)z_0)^{-1}| = 1$. This equation has just one solution given by $[2\mathrm{Re}(\lambda(A))]^{-1}$, so that we may conclude that the convergence region of the inner iteration method in (3.7) contains the domain

$$(4.7)\qquad \mathbb{D} = \Big\{\mathbf{z}:\ z_1^2 + z_2^2 < 2z_0^2,\ z_0 := \max_{\mathrm{Re}(\lambda(A))<0} \frac{1}{2\mathrm{Re}(\lambda(A))}\ ,\ z_1 \leq 0,\ z_2 \leq 0\Big\}.$$

Suppose that the matrices J_1 and J_2 possess the spectral radius $\rho(J_1)$ and $\rho(J_2)$. Then the convergence condition becomes $h^4(\rho^2(J_1) + \rho^2(J_2)) < 2z_0^2$. Thus, we have the convergence result:

Theorem 4.3. Let $\sigma = 2$. If A has one or more eigenvalues in the negative halfplane, then a sufficient condition for convergence of AF iteration {(3.7), B = A} is given by

$$(4.8) \qquad h < \left(\frac{2z_0^2}{\rho^2(J_1) + \rho^2(J_2)} \right)^{1/4}, \quad z_0 := \max_{\mathrm{Re}(\lambda(A))<0} \frac{1}{2\mathrm{Re}(\lambda(A))} \ . \ \blacklozenge$$

Example 4.1. We illustrate this convergence result by means of the RKN method generated by the fifth-order Radau IIA method for first-order ODE methods. From (2.2) it follows that the RKN matrix A is the square of the Radau IIA matrix, so that

$$(4.9) \qquad A = \begin{pmatrix} \frac{88 - 7\sqrt{6}}{360} & \frac{296 - 169\sqrt{6}}{1800} & \frac{-2 + 3\sqrt{6}}{225} \\ \frac{296 + 169\sqrt{6}}{1800} & \frac{88 + 7\sqrt{6}}{360} & \frac{-2 - 3\sqrt{6}}{225} \\ \frac{16 - \sqrt{6}}{36} & \frac{16 + \sqrt{6}}{36} & \frac{1}{9} \end{pmatrix}^2 \approx \begin{pmatrix} 0.022 & -0.020 & 0.010 \\ 0.177 & 0.038 & -0.007 \\ 0.318 & 0.182 & 0.000 \end{pmatrix}.$$

Its eigenvalues are given by $\lambda(A) \approx 0.0756$ and $\lambda(A) \approx -0.0078 \pm 0.0601i$. Applying Theorem 4.3 results into the convergence condition $h < 9.52\ [\rho^2(J_1) + \rho^2(J_2)]^{-1/4}$. ♦

When A has one or more eigenvalues in the left halfplane, one may wonder whether the fixed point iteration process might be a better approach than the AF process. To answer this question, we should consider the fixed point error equation. By observing that using fixed point iteration for solving the Newton systems in (3.6a) yields an inner-outer iteration process of the form (3.7) with B = O, i.e. $\Pi = I$, the inner amplification matrix Z reduces to

$$(4.10) \qquad Z = I - M = A \otimes h^2 J.$$

This relation shows that fixed point iteration converges if $h < [\rho(J)\rho(A)]^{-1/2}$. A comparison with (4.8) yields the theorem

Theorem 4.4. Let $\sigma = 2$. If A has one or more eigenvalues in the negative halfplane, then the interval of convergent stepsizes h of AF iteration {(3.7), B = A} is larger than that of fixed point iteration {(3.7), B = O} if

$$(4.11) \qquad \frac{\rho^2(J_1) + \rho^2(J_2)}{\rho^2(J_1 + J_2)} < \frac{1}{2} \min_{\mathrm{Re}(\lambda(A))<0} \left(\frac{\rho(A)}{\mathrm{Re}(\lambda(A))} \right)^2 \ . \ \blacklozenge$$

For example, if we use a splitting according to dimensions in the two-dimensional wave equation, then $\rho(J_1) = \rho(J_2) = \rho(J_1 + J_2)/2$, so that the lefthand side of (4.11) becomes 1/2. Hence, (4.11) is always satisfied.

There are of course other aspects that should be taken into account. AF iteration needs LU decompositions and forward-backward substitutions. On the other hand, the amplification factor is much better for AF iteration. In order to appreciate the damping of the initial error $\mathbf{Y}^{(0)} - \mathbf{Y}_{n+1}$ by the two iteration methods, we compare the amplification factor (4.5) with the amplification factor associated with (4.10). For the AF method, the largest amplification factors occur on the line $z_1 = z_2 = z/2$, so that along this line their magnitudes are respectively given by

$$\zeta_{AF} = \max_{Re(\lambda(A))<0} \frac{|\lambda(A)|^2 z^2}{4 - 4Re(\lambda(A))\, z + |\lambda(A)|^2 z^2} \quad , \quad \zeta_{FP} = |\rho(A)z| , \quad z := h^2\lambda(J).$$

An important aspect is that ζ_{AF} increases only slightly beyond 1, so that using too large stepsizes never causes a violent divergence behaviour as would be the case when fixed point iteration is applied. In fact, ζ_{AF} will never exceed the value $\left(1 - [Re(\lambda(A)).|\lambda(A)|^{-1}]^2\right)^{-1}$. For example, in the case of the fifth-order Radau IIA based method (4.9), this maximal value is about 1.017.

4.3.2. Three splitting terms. For three splitting terms ($\sigma = 3$) we can obtain a spectrum condition on A by using the following lemma (for a proof see [2]):

Lemma 4.1. Let $\mathbf{w} := (w_1, w_2, w_3)$ and define the functions $p(\mathbf{w}) := (1 - w_1)(1 - w_2)(1 - w_3)$ and $m(\mathbf{w}) := 1 - \mathbf{e}^T\mathbf{w}$, where w_j are complex variables. Then, in the region $\{\mathbf{w}: 3\pi/4 \le \arg(w_j) \le 5\pi/4\}$, the function $1 - p^{-1}(\mathbf{w})m(\mathbf{w})$ assumes values within the unit circle. ♦

From (4.4) it follows that $\zeta(\mathbf{z}) = 1 - p^{-1}(\lambda(A)\mathbf{z})\, m(\lambda(A)\mathbf{z})$. Applying Lemma 4.1 with $w_j = \lambda(A)z_j$, we see that $\zeta(\mathbf{z})$ assumes values within the unit circle in the region $\{\mathbf{z}: 3\pi/4 \le \arg(\lambda(A)z_j) \le 5\pi/4\}$. Thus, we have the result:

Theorem 4.5. If A has eigenvalues $\lambda(A)$ with $|\arg(\lambda(A))| \le \pi/4$, then AF iteration {(3.7), B = A} is A(0)-convergent for $\sigma = 3$. ♦

Corollary 4.2. If the generating GLM $(\tilde{\mathbf{a}}, \tilde{R}, \tilde{S}, \tilde{T})$ in the indirect GLM (2.2) satisfies $|\arg(\lambda(\tilde{T}))| \le \pi/8$, then AF iteration {(3.7), $A = B = \tilde{T}^2$} is A(0)-convergent for $\sigma = 3$. ♦

Hence, AF iteration applied to the examples (2.3), (2.4) and (2.5) is also A(0)-convergent for $\sigma = 3$. However, this is not the case for the RKN methods generated by the Radau IIA, Lobatto IIIA and Gauss methods, because they all have $|\arg(\lambda(\tilde{A}))| > \pi/8$.

If $|\arg(\lambda(A))| > \pi/4$, then the convergence region is finite and the region of divergence is a sort of hyperboloid. In order to get some idea of the region of convergence, we plotted in Figure 2 for (2.4) the convergence boundaries in the (z_1,z_2)-plane for a few values of z_3.

By virtue of the symmetry with respect to z_j, the convergence region contains the domain (cf. (4.7))

$$\mathbb{D} := \bigcap_{\lambda(A)} \left\{ \mathbf{z}: z_1^2 + z_2^2 + z_3^2 < 3z_0^2,\ z_j \le 0,\ j = 1, 2, 3 \right\},$$

where z_0 is the negative root of smallest magnitude of the equation $|1 - \pi^{-1}(z_0\mathbf{e})\mu(z_0\mathbf{e})| = 1$, that is, of the equation $|\pi(z_0\mathbf{e})|^2 - |\pi(z_0\mathbf{e}) - \mu(z_0\mathbf{e})|^2 = 0$. Let us write $\lambda(A) = r\exp(i\alpha)$, and define $q := |z_0 r|$. Then, it can be shown that this equation yields the following relation between q and α:

$$(4.12) \quad [1 + 3q^2 - 6q^4] + 6q[1 + 2q^2]\cos(\alpha) + 4q^2[3 - 2q + 3q^2]\cos^2(\alpha) = 0,\ q \ge 0,\ \alpha \ge \pi/4.$$

The value of q defined by this relation equals ∞ at $\alpha = \pi/4$, then rapidly decreases to ≈ 0.85 at $\alpha = \pi/2$, and slowly decreases to ≈ 0.33 at $\alpha = \pi$. The relation (4.12) leads to the following analogue of Theorem 4.3

Theorem 4.6. Let $q = q(\alpha)$ be the defined by (4.12). Then, for $\sigma = 3$ a sufficient condition for convergence of AF iteration {(3.7), B = A} is given by

$$(4.13) \qquad h < \left(\frac{3z_0^2}{\rho^2(J_1) + \rho^2(J_2) + \rho^2(J_3)} \right)^{1/4},\quad z_0 := -\min_{\lambda(A)} \frac{q(\arg(A))}{|\lambda(A)|} \quad . \blacklozenge$$

4.4. Matrices B ≠ A

In this section, we investigate whether the severe conditions on the spectrum of the matrix A to achieve A(0)-convergence derived in the preceding Section 4.3 can be relaxed by choosing B ≠ A. Some insight can be obtained by looking at the behaviour of the amplification matrix $Z(\mathbf{z})$ at infinity. We respectively consider $Z(\mathbf{z})$ in the cases where $z_i \to \infty$ and $z_j = 0$ for $j \ne i$, and in the case where all components z_i tend to infinity. This yields, respectively,

$$(4.14) \qquad \begin{aligned} &Z(\mathbf{z}) \approx I - B^{-1}A + z_i^{-1}B^{-1}(I - B^{-1}A) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{as } z_i \to \infty,\ i = 1, \ldots, \sigma. \\ &Z(\mathbf{z}) \approx I - \delta B^{-\sigma}A,\quad \delta := \frac{(-1)^{\sigma+1}(\mathbf{e}^T\mathbf{z})}{z_1 \cdot \ldots \cdot z_\sigma} \end{aligned}$$

Since $z_i < 0$ and $\delta > 0$ we easily derive from (4.14) the following result:

Theorem 4.7. For $\sigma \ge 2$, the conditions $|\lambda(I - B^{-1}A)| \le 1$ and $\mathrm{Re}(\lambda(B^{-\sigma}A)) \ge 0$ are necessary for the A(0)-convergence of AF iteration. ♦

This theorem provides a guide line for choosing the matrix B.

Example 4.2. Consider the method (2.4). The eigenvalues of the matrix A are given by $\lambda(A) \approx 0.0556 \pm 0.1571\ i$, so that $|\arg(\lambda(A))| \approx 0.39\ \pi$. If we would have chosen B = A, then the first necessary A(0)-convergence condition of Theorem 4.7 is trivially satisfied. However, since

$|\arg(\lambda(B^{-\sigma}A))| = |\arg(\lambda(A^{1-\sigma}))| = (\sigma-1)\,|\arg(\lambda(A))| \approx 0.39\,(\sigma-1)\,\pi$, the second condition of this theorem is violated if $0.39\,(\sigma-1)\,\pi > \pi/2$, i.e. if $\sigma > 2.28$.
Now, let us choose B diagonal and such that $I - B^{-1}A$ has two zero eigenvalues, so that the first condition of Theorem 3.6 is satisfied. This leads to

$$B = \frac{1}{18}\begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}. \tag{4.15}$$

A straightforward calculation reveals that

$$B^{-\sigma}A = 18^{\sigma-1}\begin{pmatrix} 2 & -1 \\ 9^{1-\sigma} & 0 \end{pmatrix}, \quad \lambda(B^{-\sigma}A) = 18^{\sigma-1}\left(1 \pm \sqrt{1 - 9^{1-\sigma}}\right), \tag{4.16}$$

so that the second condition of Theorem 4.7 is also satisfied, irrespective the value of σ. For $\sigma = 2$ and $\sigma = 3$, we checked the A(0)-convergence in the case where B is defined by (4.15) and verified that in both cases we have A(0)-convergence. ♦

5. Fixed numbers of inner and outer iterations

If the implicit relations (3.2) are iterated until convergence, then we may rely on the order of accuracy and the stability of the underlying GLM (2.1). However, in actual computation, it is often more efficient if we do *not* iterate the outer and inner iteration process until convergence. Consequently, the order of accuracy and the stability properties of the resulting integration scheme will not be identical to those of the underlying integration method. On the other hand, there is no need for convergence of the AF iteration process.

5.1. Order of accuracy

Let us consider the order of accuracy of the step values produced by the iterated method for fixed m and r (we recall that a step value is an external stage value corresponding to a step point t_{n+1}). Let $\mathbf{u}_{n+1,i}$ be a step value in the underlying method (2.1) and let $\mathbf{u}_{n+1,i}^{(m,r)}$ be the approximation after m outer and r inner iterations. If $\mathbf{u}_{n+1,i}$ has local error of order p+1, then

$$\mathbf{u}_{n+1,i}^{(m,r)} - \mathbf{y}(t_n + h) = \mathbf{u}_{n+1,i}^{(m,r)} - \mathbf{u}_{n+1,i} + \mathbf{u}_{n+1,i} - \mathbf{y}(t_{n+1}) = \mathbf{u}_{n+1,i}^{(m,r)} - \mathbf{u}_{n+1,i} + O(h^{p+1}), \tag{5.1}$$

where $\mathbf{y}(t_{n+1})$ denotes the locally exact solution. By observing that no iteration errors are introduced in the computation of the explicit stages, we can derive the order in h of $\mathbf{u}_{n+1,i}^{(m,r)} - \mathbf{u}_{n+1,i}$ by using the iteration error estimate (4.2'). Let the predictor for the implicit stage values have local error of order q+1, i.e. $\varepsilon^{(0,r)} = O(h^{q+1})$. Then (5.1) and (4.2') yield

$$\| \mathbf{u}_{n+1,i}^{(m,r)} - \mathbf{y}(t_n + h)\| = h^{q+1}\left(O(h^{\theta r}) + O(h^{u+2})\right)^m + O(h^{p+1}). \tag{5.2}$$

Thus, the maximal order of accuracy is reached if $m \geq (p - q) / \min\{\theta r, u+2\}$.

5.2. Stability

In order to see the effect of the number of iterations on the stability, we apply the integration process to the stability test equation $\mathbf{y}' = J\mathbf{y}$. We shall confine our considerations to the case where S and T have the structure as specified in (3.4), so that the GLM can be written in the form (3.5b).
Since the test equation is linear, we may set $\mathbf{G}_n = \mathbf{0}$ in (4.1). From (4.2) and (3.5b) it follows

$$\mathbf{Y}^{(m,r)} - \mathbf{Y}_{n+1} = Z^{rm}\big(\mathbf{Y}^{(0,r)} - \mathbf{Y}_{n+1}\big), \quad \mathbf{Y}_{n+1} = M^{-1}\big((R_2\otimes I)\mathbf{U}_n + (S_{22}\otimes h^2J)\mathbf{Y}_n\big).$$

Let the predictor for the outer iteration process be given by $\mathbf{Y}^{(0,r)} = P\mathbf{U}_n$. Then,

$$(5.3)\quad \mathbf{Y}^{(m,r)} = Z^{rm}P\mathbf{U}_n + \big(I - Z^{rm}\big)M^{-1}\big((R_2\otimes I)\mathbf{U}_n + (S_{22}\otimes h^2J)\mathbf{Y}_n\big).$$

By identifying $\mathbf{Y}_{n+1}$ with $\mathbf{Y}^{(m,r)}$ it follows from (3.5b) that for the stability test equation

$$\mathbf{X}_{n+1} = (R_1\otimes I)\mathbf{U}_n + (S_{12}\otimes h^2J)\mathbf{Y}_n,$$

$$(5.4)\quad \mathbf{Y}_{n+1} = Z^{rm}P\mathbf{U}_n + \big(I - Z^{rm}\big)M^{-1}[(R_2\otimes I)\mathbf{U}_n + (S_{22}\otimes h^2J)\mathbf{Y}_n],$$

$$\mathbf{Z}_{n+1} - (T_{32}A^{-1}\otimes I)\mathbf{Y}_{n+1} = \big((R_3 - T_{32}A^{-1}R_2)\otimes I\big)\mathbf{U}_n + \big((S_{32} - T_{32}A^{-1}S_{22})\otimes h^2J\big)\mathbf{Y}_n.$$

Thus, we obtain a relation of the type $\mathbf{U}_{n+1} = \Sigma_{mr}\mathbf{U}_n$, where Σ_{mr} is a matrix defined by (5.4) and which depends on the matrices h^2J_i. Its eigenvalues are given by the eigenvalues of the matrix $\Sigma_{mr}(\mathbf{z})$, where $\Sigma_{mr}(\mathbf{z})$ is defined in the same way as the matrices $M(\mathbf{z})$, $\Pi(\mathbf{z})$ and $Z(\mathbf{z})$ in (4.3). Next we observe that due to possible internal stages, the matrix $\Sigma_{mr}(\mathbf{z})$ may contain a number of zero columns. As a consequence, the corresponding components of $\mathbf{U}_{n+1}$ do not play a role in the propagation of perturbations through the steps. Let all ith columns of $\Sigma_{mr}(\mathbf{z})$ with $i \in \mathbb{I}$ be a zero column, and let $\tilde{\Sigma}_{mr}(\mathbf{z})$ denote the matrix obtained by removing all ith columns and ith rows from $\Sigma_{mr}(\mathbf{z})$ for $i \in \mathbb{I}$. Then, we have stability if the *stability matrix* $\tilde{\Sigma}_{mr}(\mathbf{z})$ has its eigenvalues on the unit disk. The region of stability is defined by the region in the $\mathbf{z}$-plane where $\tilde{\Sigma}_{mr}(\mathbf{z})$ has its eigenvalues within the unit circle (cf. the definition of the region of convergence in Section 3). Again assuming that the eigenvalues of the 'partial' Jacobians J_i are on the nonpositive real axis, we shall call the integration method *A(0)-stable* if the region of stability contains the region $\{\mathbf{z}: z_i \le 0\}$.
We illustrate the above procedure by deriving the stability matrix for iterated RKN methods with step point value $\mathbf{y}_{n+1} = (\mathbf{e}_s^T\otimes I)\mathbf{Y}_{n+1}$ and with only one explicit derivative stage value $h\mathbf{y}'_{n+1}$, i.e. $\mathbf{U}_{n+1} = \big(\mathbf{Y}_{n+1}^T, h\mathbf{y}'^T_{n+1}\big)^T$. Using the 'last step value' predictor $\mathbf{Y}^{(0,r)} = P\mathbf{U}_n = (\mathbf{e}\mathbf{e}_s^T\otimes I)\mathbf{Y}_n$, we have

$$(5.5)\quad R = \begin{pmatrix} \mathbf{e}\mathbf{e}_s^T & \mathbf{c} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad S = O, \quad T = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{d}^T & 0 \end{pmatrix},$$

where $\mathbf{c}$ and $\mathbf{d}$ are s-dimensional vectors.The equations (5.4) take the form

$$\mathbf{Y}_{n+1} = Z^{rm}(\mathbf{e}\mathbf{e}_s^T\otimes I)\mathbf{Y}_n + \big(I - Z^{rm}\big)M^{-1}\big((\mathbf{e}\mathbf{e}_s^T\otimes I)\mathbf{Y}_n + h(\mathbf{c}\otimes I)\mathbf{y}'_n\big),$$

(5.4')
$$h\mathbf{y}'_{n+1} - (\mathbf{d}^T A^{-1}\otimes I)\mathbf{Y}_{n+1} = \big((- \mathbf{d}^T A^{-1}\mathbf{e}\mathbf{e}_s^T)\otimes I\big)\mathbf{Y}_n + h\big((1 - \mathbf{d}^T A^{-1}\mathbf{c})\otimes I\big)\mathbf{y}'_n.$$

Using $\mathbf{y}_{n+1} = (\mathbf{e}_s^T\otimes I)\mathbf{Y}_{n+1}$, we obtain

$$\mathbf{Y}_{n+1} = \Big((I - Z^{rm})M^{-1} + Z^{rm}\Big)(\mathbf{e}\otimes I)\mathbf{y}_n + h\Big(I - Z^{rm}\Big)M^{-1}(\mathbf{c}\otimes I)\mathbf{y}'_n,$$

(5.4'')
$$\mathbf{y}_{n+1} = (\mathbf{e}_s^T\otimes I)\Big((I - Z^{rm})M^{-1} + Z^{rm}\Big)(\mathbf{e}\otimes I)\mathbf{y}_n + h(\mathbf{e}_s^T\otimes I)\Big(I - Z^{rm}\Big)M^{-1}(\mathbf{c}\otimes I)\mathbf{y}'_n,$$

$$h\mathbf{y}'_{n+1} - (\mathbf{d}^T A^{-1}\otimes I)\mathbf{Y}_{n+1} = - \big(\mathbf{d}^T A^{-1}\mathbf{e}\otimes I\big)\mathbf{y}_n + h\big((1 - \mathbf{d}^T A^{-1}\mathbf{c})\otimes I\big)\mathbf{y}'_n.$$

Elimination of $\mathbf{Y}_{n+1}$ leads to the 2-by-2 stability matrix

(5.6)
$$\tilde{\Sigma}_{mr}(\mathbf{z}) = \begin{pmatrix} \mathbf{e}_s^T\big(S_{mr}(\mathbf{z}) + Z^{mr}(\mathbf{z})\big)\mathbf{e} & \mathbf{e}_s^T S_{mr}(\mathbf{z})\mathbf{c} \\ \mathbf{d}^T A^{-1}\big(S_{mr}(\mathbf{z}) + Z^{mr}(\mathbf{z}) - I\big)\mathbf{e} & 1 + \mathbf{d}^T A^{-1}\big(S_{mr}(\mathbf{z}) - I\big)\mathbf{c} \end{pmatrix},$$

where $S_{mr}(\mathbf{z}) := \big(I - Z^{mr}(\mathbf{z})\big)M^{-1}(\mathbf{z})$. It is of interest to study the behaviour of the stability matrix $\tilde{\Sigma}_{mr}(\mathbf{z})$ at infinity. We consider $\tilde{\Sigma}_{mr}(\mathbf{z})$ in the cases where $z_i \to \infty$ and $z_j = 0$ for $j \neq i$, and in the case where all components z_i tend to infinity. From the relations (4.14) and

$$M^{-1}(\mathbf{z}) \approx z_i^{-1}A^{-1},\ S_{mr}(\mathbf{z}) \approx O(z_i^{-1}) \quad \text{as} \quad z_i \to \infty, \quad i = 1, \ldots, \sigma,$$
$$M^{-1}(\mathbf{z}) \approx \varepsilon A^{-1}, \quad S_{mr}(\mathbf{z}) \approx O(\delta\varepsilon) \quad \text{as} \quad \varepsilon, \delta \to 0,$$

where $\varepsilon := - (\mathbf{e}^T\mathbf{z})^{-1}$ and δ is defined in (4.14), it follows that the two eigenvalues of $\tilde{\Sigma}_{mr}(\mathbf{z})$ approach the values $\{\mathbf{e}_s^T(I - B^{-1}A)^{mr}\mathbf{e},\ 1 - \mathbf{d}^T A^{-1}\mathbf{c}\}$ and $\{1 - mr\delta(\mathbf{e}_s^T B^{-\sigma}A\mathbf{e}),\ 1 - \mathbf{d}^T A^{-1}\mathbf{c}\}$, respectively. Since $|\ 1 - \mathbf{d}^T A^{-1}\mathbf{c}\ | \le 1$ is also needed for the A(0)-stability of the underlying RKN method, we have:

Theorem 5.1. Let the underlying GLM (3.5) be an A(0)-stable RKN method defined by (5.5) and let the initial iterate for AF iteration be defined by $\mathbf{Y}^{(0,r)} = (\mathbf{e}\mathbf{e}_s^T\otimes I)\mathbf{Y}_n$. Then, after m outer and r inner iterations, the two conditions $|\ \mathbf{e}_s^T(I - B^{-1}A)^{mr}\mathbf{e}\ | \le 1$ and $\mathbf{e}_s^T B^{-\sigma}A\mathbf{e} \ge 0$ are necessary for the A(0)-stability of the iterated RKN method. ♦

Example 5.1. In the case of the A(0)-stable, third-order Radau based RKN method (2.4), we find for B = A that $|\mathbf{e}_s^T(I - B^{-1}A)^{mr}\mathbf{e}| = 0$ for all mr, but already for $\sigma = 2$ we have $\mathbf{e}_s^T B^{-\sigma}A\mathbf{e} = \mathbf{e}_s^T A^{1-\sigma}\mathbf{e} = -14$. Hence, according to Theorem 5.1, we cannot have A(0)-stability. Figure 3 presents numerical plots for a few values of mr.
However, if we define B by (4.15), then the first condition is still satisfied because the spectral radius of $I - B^{-1}A$ vanishes and hence $(I - B^{-1}A)^{mr}$ vanishes for $mr \ge 2$ (s-by-s matrices M with only zero eigenvalues have the property that $M^n = O$ for $n \ge s$). Furthermore, it follows from (4.16) that $\mathbf{e}_s^T B^{-\sigma}A\mathbf{e} = 2^{\sigma-1}$, so that the second necessary A(0)-stability condition of Theorem 5.1 is also satisfied. Numerical plots for $\sigma = 2$ show A(0)-stability for all values of mr. ♦

6. Concluding remarks

In this paper, we have analysed an outer-inner iteration method based on modified Newton and approximate factorization for solving the implicit relations occurring in General Linear Methods (GLMs) for second-order ODEs originating from multi-dimensional wave equations. The implicit relations are characterized by a matrix A, the iteration method by a matrix B.

Convergence conditions can be expressed in terms of spectral properties of the matrices A and B. Table 6.1a summarizes the main convergence results for second-order equations as derived in the present paper and Table 6.1b compares them with the A-convergence results for first-order equations derived in [2]. In these tables, A_{R3} indicates the Butcher matrix of the 3rd-order Radau IIA method for first-order ODEs, and $\tilde{A}$ and $\tilde{B}$ refer to the matrices used in AF iteration for first-order ODEs.

The stability conditions for the AF iterated methods depend on the product mr of the number of outer and inner iterations. Easy to check conditions that are necessary for A(0)-stability have been derived for a family of Runge-Kutta-Nyström (RKN) methods. The tables 6.2 list the main results.

Table 6.1a. Second-order ODEs
Cases of A(0)-convergence

σ	$B = A$	$\rho(I - B^{-1}A) = 0$
2	$Re(\lambda(A)) \geq 0$	$A = A^2_{R3}$
3	$\lvert arg(\lambda(A))\rvert \leq \pi/4$	$A = A^2_{R3}$
≥ 4	$\lambda(A) \geq 0$	

Table 6.1b. First-order ODEs
Cases of A-convergence

σ	$\tilde{B} = \tilde{A}$	$\rho(I - \tilde{B}^{-1}\tilde{A}) = 0$
2	$\lambda(\tilde{A}) \geq 0$	$\tilde{A} = A_{R3}$

Table 6.2a. Second-order ODEs
Cases of A(0)-stability

σ	$B = A$	$\rho(I - B^{-1}A) = 0$
2	$A = A^2_{R3}$, $mr = \infty$	$A = A^2_{R3}$, $mr \geq 1$

Table 6.2b. First-order ODEs
Cases of A-stability

σ	$\tilde{B} = \tilde{A}$	$\rho(I - \tilde{B}^{-1}\tilde{A}) = 0$
2	$\tilde{A} = A_{R3}$, $mr \geq 1$	

References

[1] Butcher, J.C. [1987]: The Numerical Analysis of Ordinary Differential Equations, Runge-Kutta and General Linear Methods, Wiley.

[2] Eichler-Liebenow, C., Houwen, P.J. van der & Sommeijer, B.P. [1997]: Analysis of approximate factorization in iteration methods, to appear in APNUM.

[3] Hairer, E. [1979]: Unconditionally stable methods for second order differential equations, Numer. Math. 32, 373-379.

[4] Hairer, E. & Wanner, G. [1991]: Solving ordinary differential equations, Vol. II. Stiff and differential-algebraic problems, Springer-Verlag, Berlin.

[5] Houwen, P.J. van der, Sommeijer, B.P. & Kok, J. [1997]: The iterative solution of fully implicit discretizations of three-dimensional transport models, APNUM.

[6] Shampine, L.F. [1980]: Implementation of implicit formulas for the solution of ODEs, SIAM J. Sci. Stat. Comput. 1, 103-118.

[7] Sharp, P.W., Fine, J.H. & Burrage, K. [1990]: Two-stage and three-stage diagonally implicit Runge-Kutta-Nyström methods of orders three and four, IMA J. Numer. Anal. 10, 489-504.